

**A STUDY OF SCHOOL FACTORS
AFFECTING ACHIEVEMENT**

- AN IDENTIFICATION APPROACH

K.R. PRASAD

ADVISORY COMMITTEE

DR. R.H. DAVE (CHAIRMAN)

DR. R.G. MISRA (MEMBER)

SH. I.S. CHAUDHRI (MEMBER-SUPERVISOR)

**SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE ASSOCIATESHIP OF THE NATIONAL INSTITUTE OF EDUCATION,
NEW DELHI**

1967-68

ACKNOWLEDGEMENT

The Investigator expresses his gratitude to the Chairman and Members of the Advisory Committee for their kind cooperation and constant encouragement in completing the study.

Thanks are due to his friends for their valuable suggestions and active co-operation.

The Investigator acknowledges with thanks the Department of Field Services (N.I.E.), New Delhi, for making the data available.

- PRASAD

CONTENTS

CHAPTER	PAGES
I The Problem	1-4
II Approach	5-7
III Procedure	8-14
IV METHODOLOGY	15-19
V Results and Interpretation	20-29
VI Limitations and Suggestions	30-31
VII Summary	32-34
 REFERENCES	 35

APPENDIX

I Achievement Quotients (A.Q) of the Schools	I
IIA Formulae for getting the variables	II
IIB Measurements on Factors	III
III Differences in Mean Values of the Factors	IV
IV Pooled Variances and Co-variances of Factors	V
V Pivotal Condensation Method for obtaining the discriminant function Co-efficients	VI-VII
VI 'Z' Values for High and Low Achieving Schools.	VIII

Chapter I - THE PROBLEM

Education in a modern society can be very much likened to 'industry'. While planning for the expansion of this industry, just as in any other industry, it is essential to take into consideration the internal constraints and the real resources available. The internal constraints in this case are 'the availability of competent teachers, of physical plant and finance'. In a developing economy like ours, where the total expenditure on education hardly represents 2.9% of the National Income (at the end of the third Five Year Plan)¹, these built-in constraints make it almost impossible to provide all the facilities. Hence the problem arises as how to spend the limited resources on these 'needed facilities' in an effective manner so that the return on investment is maximized.

Taking school as a unit, the return on investment is represented by the achievement of the school. The achievement of the school, of course, depends on conditions inside as well as outside the school; variations in individual abilities apart. Conditions inside the school are represented by 'teaching staff, total number of hours for instruction, class size, expenditure per pupil, library facilities, equipment etc' and the conditions outside the school are represented by social and environmental factors. The former are well within the control of the educational authorities. As far as the latter are concerned, the authorities are helpless, in a way. Until and unless the general condition of the economy is improved, for which concerted efforts will

1. Report of the Education Commission (1964-66), Ministry of Education, Government of India, P 465

have to be made, not much progress can be achieved in improving these conditions. Under the circumstances the Educational or School authorities will have to take great care in improving the conditions inside the school, with the limited resources available, in order to raise the standards of achievement as well as to expand the facilities to meet the growing needs of the society without diluting the standards. In attempting this, the natural questions that arise are - what exactly are the conditions inside the school that are likely to affect its achievement/ and,

once these factors are arrived at, what will be their relative weightages or importance as far as the achievement is concerned. The present investigation is undertaken in search of solution to these problems, with the belief that such a solution might ultimately help in determining the priorities and thus helping in improving the achievement. The investigation is limited only to certain factors which are felt important, due to lack of time, resources and availability of data on the part of the investigator.

Statement:

To sumup, the problem may be stated as 'A study of school factors affecting achievement' consisting of (i) identification of the factors affecting the achievement of the schools and (ii) determining the relative weightages of these factors in relation to achievement.

Related Studies:

Although many studies are conducted on the general and

specific abilities of individuals and environmental and social factors affecting achievement, in and outside India, yet the same cannot be claimed about the 'conditions inside the school'.

Recently N.P. Pillai¹ has made 'an investigation into the organizational and administrative factors which affect the achievement of pupils in Secondary Schools', in Trivandrum district of Kerala, with the major objectives of finding (i) how adequately schools are equipped for instruction and how far the proper atmosphere and incentives are provided for securing achievement, (ii) how far achievement depends on factors other than instruction (iii) the relative achievement of pupils in well equipped and ill equipped schools and (iv) suggesting ways and means of improving standards. For this purpose data are collected on conditions inside as well as outside the school and are analysed with the help of analysis of variance technique. (The major findings of the study are (i) although the maximum influence is exerted by the environmental factors, the effect of the non-instructional factors is also considerable and (ii) because of the presence of interaction between the various factors, good instruction becomes more effective when the school has such facilities as a good laboratory, literary and other clubs, audio-visual aids etc. One of the most important findings of this study is when the majority of the pupils come from an environment of poverty results cannot be improved even with good teaching and proper facilities

1. Pillai, N.P., An Investigation into the organizational and Administrative Factors which affect achievement of pupils in Secondary Schools, Indian Educational Review, NCERT, New Delhi, Vol 1, No.1 July 1966 Pp.17-33

in the school. However this needs confirmation with the help of a wider study, as the sample taken up in this study is very small i.e. 24 schools.

Rasik Shah¹ has conducted 'a study of the relationship of certain educational facilities to the scholastic achievement in the secondary schools'. The facilities considered are size of the class, pupil teacher ratio, percentage of trained teachers and recurring expenditure per pupil. The problem was investigated separately for government and non-government schools as well as for boys' and girls' schools. A comparative study of two states viz Gujarat and Kerala was attempted to see if any regional variation existed in this regard. The results showed that although there is considerable variation with regard to most of the educational facilities provided in the States of Kerala and Gujarat, yet there is not much variation in government and non-government schools and boys' and girls' schools. The correlation coefficients and the multiple correlation coefficient between these facilities and the results of scholastic achievement obtained, are generally low and as such no definite conclusions could be drawn from the study. This could be due to limitations of 'poor criterion variable, inadequate sample and many other intervening variables which could not be controlled'.

1. Shah, Rasik, A Study of the Relationship of certain Educational Facilities to the Scholastic Achievement in the Secondary Schools, unpublished dissertation, Research Methodology Course, 1966-67, NIE, New Delhi .

Chapter II - THE APPROACH

The achievement of a school, as stated earlier, is affected by many factors. These factors taken individually may or may not provide useful information regarding the achievement, for no single factor need operate by itself. Some factors may show significant differences in achievement, but as these are likely to be highly correlated, the evidence they provide cannot be treated as independent. Due to the same reason other factors, which by themselves provide no means of discrimination with regard to achievement, may in conjunction with the rest aid considerably. That is to say the interaction between factors may have to do much with achievement.

Taking the 'school' as an independent unit and studying the relationship between the school factors and achievement, certainly helps in arriving at some of the factors and in knowing their relative weightages. However, the main concern of this investigation is to know the relative weightages of the relevant factors, so that the general standards of achievement may be improved and the educational facilities may be increased, by using the resources in an effective manner. In other words we are concerned with the overall improvement rather than the schools individually. As such if we can divide them into a dichotomy of high and low achieving schools, the factors are likely to provide useful additional information. This dichotomy can be arrived at either by fixing a standard of achievement or with the help of a criterion variable.

Once the schools are divided into the high and low achieving groups, our problem reduces to the identification of the factors and determining their relative weightages such

that the distance between the two groups is maximized. Similar problems arose in many other fields. In statistical literature, such a problem is referred to as one of classification or discrimination and which, of late, has come to be known as that of identification. In order to solve this problem, Fisher¹ has introduced a device which closely resembles the technique of multiple regression in some ways. This may be used for this study also in a useful manner.

The device consists of determining a new variable Z as a linear combination of the known variables X_1, X_2, \dots, X_p , in such a way that the distance between the two groups is maximized. If $Z = l_1 X_1 + \dots + l_p X_p$, this amounts to determining 'a set of adjustable coefficients (l_1, \dots, l_p) in such a way as to maximize the ratio of the square of one chosen component to the sum of squares of a set of other components in an analysis of variance.' This helps in identifying whether a particular school is high achieving one or low achieving one by knowing the measurements on the variables X_1, \dots, X_p . Also only when that particular linear function, usually termed as discriminant function, is determined which, better than any other, discriminates between the two groups of schools, 'can we recognise that some measurements are useless, while others are of real evidential value'.

This device has been of much use in a number of problems ranging 'from anthropometric classification of skulls to the

1. Fisher, R.A., Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh and London, 1958 Pp 285-89

choice of the speciality salesmen'. However, the most common use will perhaps be found in the handling of taxonomic data. Fairfield Smith¹ has used this type of analysis, to determine how the different observable characters of plant progenies should be combined in selecting for any particular end. Barnard² has used 'the regressions of the means of certain measurements of Egyptian Skulls on the approximate date of burial, to ascertain what linear function of the cranial measurements obtainable shows the most distinct change with time'. Fisher himself has used this technique on two species of plants '*Iris versicolor*' and '*Iris setosa*' by taking four characters.

-
1. As quoted by Fisher in his book Statistical Methods for Research Workers P 287
 2. ibid P 287

Chapter III - THE PROCEDURE

Data and Sample:

The data collected by the Department of Field Services (National Institute of Education), New Delhi is used for analysis purposes. The Department has collected various types of data, for five years 1958-63, on the secondary schools in India in connection with NIE - HEW Project 001 'A Survey of Secondary Schools in India'. For the purpose of the present study, only the schools in Delhi are considered and the data for the year 1962-63 forms the base, because of its relative completeness. This is due to the limitations of time and the availability of data.

The sample consists of 10% of the total number of Higher Secondary Schools in Delhi (i.e. 35 schools) selected in a stratified random manner. The stratification is done on the basis of the following variables - (1) Government and private (2) Boys and Girls. (Two other variables viz Urban and Rural and High School and Higher Secondary School are also taken into account for general survey, but these are not applicable in case of Delhi)

If, in a particular stratum, the number of schools are too low, more number of schools than the 10% are taken in order to get substantial data regarding the same. As such the sample can be treated as representative, although the size of the sample is small for carrying out this analysis.

'Dichotomization' of Schools:

The scholastic achievement is taken as the achievement of the school, for practical purposes. In order to dichotomize

the schools into high and low achieving groups, the results of the schools at Higher Secondary Examination, 1962-63 are taken into consideration. Instead of fixing an arbitrary pass percentage for dividing the schools into the two groups, it is felt that it would be better if we can arrive at a cut-off point, at which the two groups are discriminated the most. The need is also felt for considering the quality, for which purpose the divisions awarded at the examination viz I, II, III divisions are taken into account.

The following procedure is adopted to arrive at the cut-off point, taking into account the quality also. "Average marks of those placed in Ist, 2nd, 3rd divisions and those failing are broadly speaking, 65, 52, 39 and 19.5 respectively. These are in the ration of 10, 8, 6 and 3"¹. 'Achievement Quotients' (A.Q) are calculated by giving these weights to the I, II & III divisions. The failures are not taken into account, thereby fixing minimum standard for achievement. If p_1 is the proportion of students passing in I division out of the total number of candidates appeared in a particular school, p_2 is the proportion in II division and p_3 is the proportion in

III division, then the achievement quotient for that school is given by, $A.Q = 10p_1 + 8p_2 + 6p_3$. Thus the maximum possible achievement for a school is ten and the minimum is zero. After calculating the achievement quotients (vide appendix I) for the schools in the sample (which is a representative one), they are arranged in descending order and differences in successive schools are calculated. The cut-off point is taken at that where this difference is maximum. In this case it has turned out to be between 3.552 and 4.312 A.Q's. In terms of pass percentages alone (ie. not taking division into consideration), this has coincided with 55% passes. It is interesting to note the coincidence between this and the cut-off point which is generally used to distinguish high and low achievers, by the university authorities and employers.

Taking this as cut-off range, 12 schools are in low achieving group and 16 are in the high achieving group. Seven schools are left out as they have not sent any students to this examination in the year 1963. In the high achieving group also only 12 schools are taken for 'it is profitable to divide the samples equally between the two populations'.¹ The four left out schools are the border cases.

1. Rao, Radhakrishna C., Advanced Statistical Methods in Biometric Research, John Wiley & Sons, Inc. New York. 1952
p. 256

The investigator is aware of the limitations of this procedure. Better methods could have been formulated but for the limitation of availability of data. e.g. Achievement quotients could have been calculated on the basis of say, five years' results and the schools could have been divided into two groups on the basis of these. This would have resulted in getting more stable groups. However this couldn't be done as the data is available, for some schools, only for the year 1963.

Factors:

All the relevant factors are enumerated to start with. This is done by taking into consideration some of the studies in this area and also the general idea regarding the factors. e.g. The Education Commission report (1964-66 consists of comments on a number of factors. However only the following factors are actively considered, for regarding others, such as co-curricular activities, built-area etc., the information is either not available at all or incomplete -

1. Total number of hours for instruction in the academic year.
2. Size of the section in the school leaving class (i.e. Class XI)
3. Percentage of hostlers

4. Book-value of equipment per pupil
5. Percentage of trained teachers
6. Pupil-teacher ratio in secondary classes
7. Annual recurring expenditure per pupil
8. Total number of internal periodical tests
9. Number of hours of homework per day
10. Number of books in the library per pupil.

However, not all of them could be taken up, for some of the variables do not seem to be providing valuable information at all, to discriminate between high and low achieving groups.

1. There is only one school providing Hostel facilities.
2. Out of 24 schools, 12 are having 100% trained teachers and almost all the schools are having more than 85% of them.
3. The number of internal periodical tests provided in most of the schools are six.
4. The number of hours of homework per day is also mostly 3 to 4 hours.

Because of these reasons, the four variables are dropped and the following six variables are selected for analysis purpose ----

X_1 : Total number of hours for instruction in the academic year.

X_2 : Size of the section in the school leaving class (i.e. class XI)

X_3 : Pupil-teacher ratio in the secondary classes.

X_4 : Book value of equipment per pupil.

X_5 : Annual recurring expenditure per pupil.

X_6 : Number of library books per pupil.

All these variables are considered to be affecting achievement. The Studies conducted on these have also established some relationships with achievement. The Education Commission (1964-66) has also made the following suggestions¹ regarding these variables.

1. The number of instructional days in the year should be increased to about 39 weeks for schools and the loss of instructional days due to examinations etc. should not exceed 21 days.
Duration of working day should be increased at the school stage!
2. Size of the class should be 40 to 45 in Secondary classes.
3. Fixing of the average pupil-teacher ratio at different stages of school education, is not enough, for it doesn't necessarily control the class size.

1. Op.cit. p.1.

4. Full utilization of laboratories, Audio-visual equipment etc.
5. 'The non-teacher costs per pupil have actually fallen, even at constant prices. That is why our primary schools are so dull and drab'. The picture is similar at higher primary stage and colleges also. However, while calculating the cost per pupil, they have taken into account the teachers' salaries also.
6. 'Only a small proportion of children have all the books at the beginning of the year Not infrequently, a proportion of students have no books at all.'

The table given in appendix II-A gives the formulae to get observations on these variables.

Delimitations of the study:

1. The study is confined to 'Delhi' city only.
2. The study is confined to the Higher Secondary classes alone in which the school leaving class is XIth
3. The study is confined to only the six factors.

Analysis:

All these factors are analysed with the help of the discrimination analysis. The Methodology and the assumptions are explained in Chapter IV - The Methodology.

Chapter IV - THE METHODOLOGY

Let X_1, \dots, X_p be p variables on which measurements are collected on a sample of N_H schools from the high achieving group and N_L schools from the low achieving group.

Let $Z = l_1 X_1 + \dots + l_p X_p$ be the linear combination of the p variables, where the coefficients l_1, \dots, l_p are to be determined in such a way that the distance between the two groups is maximum i.e. the ratio of the variances between groups to that within groups is maximum. The ratio of between to within variance of Z from the two samples is given by,

$$\frac{N_H N_L}{N_H + N_L} \frac{(l_1 d_1 + \dots + l_p d_p)^2}{\sum l_i l_j w_{ij}}$$

where d_1, \dots, d_p are the differences in means, of the p variables, of high and low achieving groups and, w_{ij} are the pooled covariances between X_i and X_j variables.

Maximizing this and observing that only ratios of the coefficients 'l' can be uniquely determined, the coefficients of the best linear function separating the two groups are obtained as solutions of the equations,

$$l_1 w_{11} + l_2 w_{12} + \dots + l_p w_{1p} = d_1$$

.....

$$l_1 w_{p1} + l_2 w_{p2} + \dots + l_p w_{pp} = d_p$$

The assumptions made are,

1. the variables x_1, \dots, x_p are continuous and normally distributed
- and 2. the variances and co-variances of variables in the two populations are homogeneous.

The actual calculations involve the following steps -

Step I: Calculation of the mean values based on N_H and N_L observations for each variable and the difference (d) between the two means (vide Appendix III).

Step II: Calculation of pooled variances and co-variances, based on $N_H + N_L - 2$ degrees of freedom for all the variables and pairs of variables respectively (vide Appendix IV).

Step III: Arranging the pooled dispersion matrix in order of magnitude, taking sign into account, starting with the variable having the maximum variance and carrying out the 'pivotal condensation method for obtaining successive ' D^2 ' values and discriminant functions $L(x)$ at each stage.¹ (vide Appendix V). The merit of this method is that we can get the values of the coefficients $L(x)$ and the corresponding ' D^2 ' at every stage - i.e. taking only x_1 , taking x_1, x_2 and so on.

Tests of Significance:

I. To find out whether the distance between the two populations is significant or not. The specific hypothesis in this case is that there are no differences in the mean values of the p variables for the two populations.

This can be tested with the help of Mahalanobis' D^2

where $D^2 = l_1 d_1 + \dots + l_p d_p$. The test¹ is,

$$F = \frac{N_H N_L (N_H + N_L - p - 1)}{p(N_H + N_L) (N_H + N_L - 2)} D^2 \quad \text{with } p \text{ and}$$

$N_H + N_L - p - 1$ degrees of freedom.

II. For testing additional information:

Let p be the number of basic variables to which are added q more variables. In order to know whether the additional information provided by the q variables regarding the distance between the two groups is significant or not, the test² is,

$$F = \frac{N_H + N_L - p - q - 1}{q} U_{q,p} \quad \text{where } U_{q,p} = \frac{1 + \frac{N_H N_L}{(N_H + N_L)(N_H + N_L - 2)} D^2_{p+q}}{1 + \frac{N_H N_L}{(N_H + N_L)(N_H + N_L - 2)} D^2_p} -$$

and D^2_{p+q} , D^2_p are the values of ' D^2 ', based on $p+q$ variables and p variables respectively, with q and $N_1 + N_2 - p - q - 1$ d.f.

1. op.cit. p.10

2. op.cit. p.10

Probability of misclassification:

The most important use of the discriminant function is to know whether a particular school belongs to the high achieving group or low achieving group, by knowing the measurements on x_1, \dots, x_p . If \bar{z}_H is the mean of the 'Z' values of the high achieving group and \bar{z}_L is that of the low achieving group and $\bar{z} = \frac{\bar{z}_H + \bar{z}_L}{2}$, then

the schools with a 'Z' value higher than \bar{z} may be classified as high achieving ones and those with a value less than \bar{z} as low achieving ones. However some schools are likely to get misclassified. Hence it is important to know the probability of such a misclassification.

$$\text{Mean square within groups} = \frac{D^2}{(N_H + N_L - 2)(N_H + N_L - p - 1)}$$

$$\text{Estimate of S.D.} = \sqrt{\text{Mean square within groups}}$$

If ' D^2 ' is the distance between groups, a deviation $D^2/2$ will cause misclassification. In terms of standard deviation, the deviation causing misclassification is

$$t = \frac{D^2/2(N_H + N_L - 2)}{\text{Est. S.D.}} \quad \text{with } N_H + N_L - p - 1 \text{ degrees of freedom.}$$

Let the probability of getting a value equal to or more than the calculated value of 't' by chance, be 'm'. For misclassification to occur, the deviation must be in one given direction, for the other half represents the occasions on which the departure of the school value from the group mean is over half the size of the group difference, but is ~~in~~ in the direction away from that which causes misclassification. Hence the probability of misclassification is given by $\frac{m}{2}$ i.e. misclassification will occur in $\frac{m}{2} \times 100$ cases out of 100.

Chapter V - THE RESULTS AND INTERPRETATION

The Coefficients:

After obtaining the dispersion matrix and the differences between the means of high and low achieving groups for each variable, the coefficients of the discriminant function obtained at each stage, with the help of pivotal condensation method are given in the following table.

Table 1a. The Coefficients $L_i(x)$ obtained at each stage and the corresponding value of ' D^2 '

Variables	x_1	x_4	x_5	x_6	x_2	x_3	Value of ' D^2 '
x) 0.002452							0.134451
x) 0.000502	0.014721						0.479903
x) -0.000042	0.012658	0.008467					0.601379
x) -0.003545	0.000255	0.007067	0.460812				1.521029
x) -0.000302	-0.001617	0.012803	0.359336	0.131731			2.451920
x) 0.002206	-0.003637	0.021702	0.290562	0.135990	0.117579		2.998229

Observing that only the ratios of the coefficients can be uniquely determined, for convenience, we can divide the above sets of coefficients by the coefficient with the least magnitude in the set. Thus we get the following sets of coefficients.

Table 1b. The coefficients $L_i(x)$ obtained at each stage after dividing each set with the coefficient having least magnitude in the set

Variables	x_1	x_4	x_5	x_6	x_2	x_3
1						
1	29.3247					
-1	301.3810	201.5952				
-13.9020		1	27.7137	1807.1059		
-1	-5.3543	42.3940	1189.8543	436.1954		
1	-1.6487	9.8377	131.7144	61.6455	53.2996	

These sets of coefficients $L_1(x)$ to $L_6(x)$ represent the coefficients of the discriminant functions arrived at, by taking into account only the variable x_1 , the variables (x_1, x_4) , ..., the variables $(x_1, x_4, x_5, x_6, x_2, x_3)$ respectively.

Testing for the significance of the discriminant function: To test for the differences in mean values, the following table gives the values of 'F' with p and $N_H + N_L - 1 - p$ degrees of freedom.

Table 2. Values of F for testing the significance of discriminant functions based on different sets of variables

No. of variables, p	VARIABLES	D ² p	F	Significance
6	$X_1, X_4, X_5, X_6, X_2, X_3$	2.998229	2.32	not significant at .05 level. Significant at .10 level - do -
5	X_1, X_4, X_5, X_6, X_2	2.451920	2.41	
4	X_1, X_4, X_5, X_6	1.521029	1.97	Not significant both at .05 and .10 levels

The calculation of further F's is not necessary as the value of F for the discriminant function based on 4 variables is not significant both at .05 and .10 levels.

From the above table it can be seen that significant discrimination is achieved between high and low achieving groups, at .10 level for both the discriminant functions based on six and five variables. Since the discriminant function based on five variables gives a significant discrimination at .10 level, it is most likely that the function based on six variables will also be significant even if X_3 does not in itself have any discriminating value. Hence it is necessary to test whether the additional information supplied by X_3 is significant or not.

Test for the significance of additional information:

For degrees of freedom 11 & 17, $F = 1.5181$ ($p=5$ and $q=1$) is not significant at .10 level. Hence the additional information supplied by X_3 is not significant.

Noting that (i) both the discriminant functions based on five and six variables are not significant at .05 level, the level which is usually considered for significance purposes,

(ii) not all the school factors which might be effecting the achievement are taken up for the study,

and (iii) the sample is quite small for this type of analysis,

it is likely that X_3 may have contributed significant additional information either with the help of a larger sample or in presence of other factors which can be included in the investigation and thus perhaps leading to a discriminant function which is significant at .05 level. A close look at the way in which the coefficients (table 1b.) are shaken with the inclusion of X_3 supports this view. Further the coefficients indicate the relative values of the variables in distinguishing the two groups and if a particular variable is not contributing significantly, the magnitude of its coefficient will be less automatically.

Hence taking all the six variables into account the final discriminant function is,

$$Z = X_1 + 61.6X_2 + 53.3X_3 - 1.6X_4 + 9.8X_5 + 131.7X_6$$

This equation shows very clearly the relative values of the six variables in distinguishing high and low achieving schools

Classifying schools into high and low achieving groups:

Let \bar{Z}_H and \bar{Z}_L be the means of the 'Z' values of high and low achieving groups respectively and $\bar{Z} = \frac{\bar{Z}_H + \bar{Z}_L}{2}$

$$\text{For this sample } \bar{Z}_H = 6889.64$$

$$\text{and } \bar{Z}_L = 5530.36$$

$$\therefore \bar{Z} = 6210.00$$

The schools with a 'Z' value greater than 6210.00 will be classified in the high achieving group and those less than 6210.00 will be classified in the low achieving group. While doing so it is possible that some schools may get misclassified. The probability of such a misclassification is given in the following section.

Probability of misclassification:

Within groups mean square = 0.008021

$$\therefore \text{Est. S.D.} = \sqrt{\text{Within groups mean square}} \\ = 0.08956$$

In terms of standard deviation, the distance causing misclassification is given by

$$t_{17} = \frac{D_6^2 / 2(N_H + N_L - 2)}{\text{Estimate of S.D.}} = 0.761$$

The probability of getting a value equal to or more than the value of t_{17} , by chance is less than .50. But for misclassification to occur the deviation must be in one given direction. Hence the probability of getting a value equal to or more than the value of t_{17} in one given direction, by chance is less than .25. Hence the misclassification will occur in less than 25 in 100 cases. The actual calculation of 'Z' values for the schools in this sample (vide appendix VI) shows that six schools, three in the high achieving group and three in the low achieving group, out of 24 schools would have got misclassified by this procedure. The probability of misclassification i.e. 25 out of 100 agrees reasonably well with this result.

INTERPRETATION:

The coefficients of the discriminant function show very clearly the relative importance of the six variables, considered in this study. The one factor that is dominating the clearly over all the other factors in distinguishing the high and low achieving schools is 'the number of library books per pupil'. The means of both the group also suggest the importance of this factor. While, on an average, the high achieving schools have about 8 books per pupil in their library, the low achieving ones have only four. This shows that the

libraries are not kept in the schools as merely store houses. The magnitude of its coefficient viz. 131.7 as against the coefficient of the factor 'total number of hours for instruction in the academic year' viz. 1 brings out the need for the immediate improvement of the library facilities.

The other two major contributors in distinguishing the high and low achieving schools are 'size of the section in school leaving class i.e. class XI' and 'pupil-teacher ratio in the secondary ~~class~~ classes' with respective coefficients as 61.6 and 53.3. The means of the 'size of the section' in high and low achieving groups are 29 and 21 respectively. The sizes are quite low and the factor is receiving a positive weight. This shows that, in the lower ranges, it is not necessary that the lower the size of the section the better the achievement.

The means of the 'pupil-teacher ratio' in the high and low achieving groups are 25 and 22 respectively. In this case also the figures are quite low and the weight received by the factor is positive. Hence again the same conclusion can be drawn that in the lower ranges it is not necessary that the lower the pupil-teacher ratio, the better the achievement.

The 'annual recurring expenditure per pupil' has also received a positive weight of 9.8. The means of the high and low achieving groups are Rs. 194.36 and

Rs. 169.00 respectively. Although the coefficient of this factor is relatively small as compared with those of the above discussed factors, the positive weight along with the means indicate that this factor does help in distinguishing high and low achieving schools. As for the priority of this factor among these six is concerned, it is of significance and also gratifying in the context of developing Indian economy where only limited funds could be allotted for the cause of education.

The average 'number of instructional ^{hours} per academic year' in the high and low achieving groups are 990 and 935 respectively. The weight received by it and the priority obtained among the six factors do not suggest the increase of number of working days per academic year and the duration of the working day on a priority basis.

The factor 'the book value of equipment per pupil' is included with a view to study the relative impact of teaching aids, laboratory facilities, audio-visual equipment etc. on the achievement of the school. The means of the high and low achieving schools, in the sample are Rs.81.84 and Rs. 51.11 respectively. Inspite of the great difference in means this factor, rather surprisingly, is receiving a negative weight, though small viz -1.6. One reason

for this can be that the schools may not be so well equipped to the extent of making any significant impact on the achievement. Another pertinent reason can be that the schools may not be making proper use of whatever equipment they possess.

It should be noted here that the study is limited to only six factors and the way in which the weightages of the factors are shaken up with every addition of a new factor, as can be seen from table 1b., suggests that the factors are highly interacting. However, consistency regarding the priorities of the factors is maintained in the last two stages that is in the case of five factors and six factors. As such there may be some more important factors affecting achievement of the schools and the relative weightages may get changed, if not the priorities, if these factors are also included. Hence the above interpretations should be viewed, taking these points into account.

It is of interest to note that the recommendations made by the Education Commission (1964-66) regarding these factors, are in tune with the findings of this study excepting in the case of 'total number of hours for instruction per academic year'. The Commission has recommended the increase in number of working days and duration of the working day as a necessity, while the

factor has not received so much importance, in this study, in relation to other factors.

As stated earlier, the sample for this study is taken only from Delhi and can be treated as representative of the population of Delhi Higher Secondary Schools. As such the results and interpretations are applicable only to Delhi, while they may be indicative of overall trend.

Chapter XI - THE LIMITATIONS AND SUGGESTIONS

Limitations:

The three major limitations of this investigation are,

- (i) the sample is rather small, although it may be taken as representative of the population of higher secondary schools in Delhi,
- (ii) the data is collected for some other purpose and it has restricted the scope of the study ,
- (iii) the criterion used for dichotomising the schools into high and low achieving groups is not standardized,

apart from other limitations of lack of time and resources due to which the study has to be restricted only to Delhi Higher Secondary schools and only to six factors.

Suggestions:

Inspite of the above limitations, one major recommendation that is brought about by this study is the need to improve the library facilities on priority basis, out of these six factors.

The other recommendations are that there is no need to stress to have the class size as low as possible beyond a certain limit because in the lower ranges it is not necessary that the lower the class size, the better the achievement. The class size of 40 in the

secondary classes suggested by the education commission may be taken as somewhere around the optimum and the same is the case with 'pupil-teacher' ratio. The next in priority comes 'annual recurring expenditure per pupil', which also includes the teachers' salaries", taken as an estimate of the cost per pupil. The study also calls for the proper use of the laboratory facilities, audio-visual equipment, teaching aids, etc., that are provided in the schools.

The issues raised by this study, on which further researches can be made, are -

1. developing a standard criterion on the basis of which the schools can be dichotomised into high and low achieving groups.
2. If C is the total amount that is to be spent on a school and if there are N schools, will it be possible to get a set of optimum weights to the factors, subjected to the total fixed cost NC. A solution can be tried for this by a similar approach.
3. Comparative studies can be made between states, as it is well-known that the conditions in different states of India are different.
- and lastly,
4. It will be useful if the study is replicated with a large sample and planned collection of dat

Chapter VII - THE SUMMARY

The investigation is taken up to identify the factors 'inside the school' affecting achievement and to determine their relative weightages in relation to achievement. The investigation is limited to Delhi Higher Secondary Schools and to only six factors because of lack of time, resources and availability of data. The six factors are,

- x_1 Total number of hours for instruction in the academic year,
- x_2 Size of the section in school leaving class,
- x_3 Pupil-teacher ratio in secondary classes,
- x_4 Book value of equipment per pupil
- x_5 Annual recurring expenditure per pupil
- x_6 Number of library books per pupil

The data employed (for the year 1962-63) is collected for some other purpose by the Department of Field Services by selecting a sample on stratified random basis. The schools are dichotomized into high and low achieving groups on the basis of the achievements of the schools taking quality also into account. The weights given to the I, II, III divisions are 10, 8, 6 respectively. 12 schools are taken in each of the high and low achieving groups.

for analysis purpose.

After calculating the dispersion matrix and the differences in means of the high and low achieving groups for each variable, the coefficients of the discriminant function are obtained by carrying out the 'Pivotal condensation method for obtaining successive Σ^2 ' values and discriminant functions ' $L(x)$ ' at each stage'. The discriminant function obtained is,

$$Z = X_1 + 61.6 X_2 + 53.3 X_3 - 1.6X_4 + 9.8 X_5 + 131.7X_6.$$

This is found to be significant at .10 level. If the schools are to be classified into one of the two groups with the help of this function, the probability of misclassification is found to be .25 i.e. 25 in 100.

The discriminant function shows very clearly the relative values of the six variables in distinguishing high and low achieving schools. Of the six variables, on a priority basis, the discriminant function brings out clearly the need for immediate improvement of the library facilities. Another fact that is brought out by this study is that, in lower ranges, it is not necessary that the lower the class size and the pupil teacher ratio, the better the achievement. Annual recurring expenditure per pupil comes next to these. Another significant finding

is that either the schools are not so well equipped (laboratory, audio-visual, teaching-aids etc) to the extent of making any significant impact on achievement or they are not utilizing the provided equipment in a proper way. In all these the study concurs with the views expressed by the Education Commission. However, the study doesn't suggest the increase of instructional days per academic year, unlike the recommendation made by the Commission.

The limitations of study are (i) small sample (ii) data collected for some other purpose is used (iii) criterion used for dichotomization of schools into high and low achievement groups is not standardized, apart from other limitations of time and resources due to which the study is restricted to Delhi Higher Secondary Schools and to only six factors.

oooo0oooo

REFERENCES

Books:

1. Anderson, T.W. (Ed.), Selected Papers in Statistics and Probability by Abraham Wald, Stanford University Press, 1957
2. Fisher, R.A., Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh and London, 1958
3. Golden, Cyril H., Methods of Statistical Analysis, John Wiley & Sons, Inc., New York, 1956
4. Mather, K., Statistical Analysis in Biology, Methuen & Co. Ltd., London, 1964
5. Rao, Radhakrishna, C., Advanced Statistical Methods in Biometric Research, John Wiley & Sons, Inc., New York, 1952
6. Rao, Radhakrishna, C., Linear Statistical Inference and its Applications, John Wiley & Sons, Inc., New York, 1966

Reports:

7. Ministry of Education, Government of India, Report of the Education Commission, 1964-66.
8. Shah, Rasik, A Study of the Relationship of Certain Educational Facilities to the Scholastic Achievement, Unpublished dissertation, Research Methodology Course, 1966-67, N.I.E., New Delhi

Journals:

9. Indian Educational Review, NCERT, New Delhi Vol. 1, No.1, July 66, pp. 17-33
10. Applied Statistics, Royal Statistical Society, London W.1, Vol. 16, No.3, 1967, pp.

APPENDICES

APPENDIX I

Achievement Quotients (A.Q.) of the Schools

Number and Proportion of pupils in division			Total (I + II + III)	Total appeared	A.Q.	Differences in A.Q.'s of successive Schools.
I	II	III				
6(.088)	43(.632)	16(.235)	65(.956)	68	7.346	
3(.143)	7(.333)	11(.524)	21(1.00)	21	7.238	0.108
-	16(.762)	4(.190)	20(.952)	21	7.236	0.002
1(.010)	50(.481)	50(.481)	101(.971)	104	6.830	0.406
-	11(.550)	8(.400)	19(.950)	20	6.800	0.030
6(.056)	60(.561)	31(.290)	97(.907)	107	6.788	0.012
-	10(.370)	15(.556)	25(.926)	27	6.296	0.492
7(.121)	21(.362)	20(.345)	48(.828)	58	6.176	0.120
-	7(.184)	27(.711)	34(.895)	38	5.738	0.438
8(.068)	39(.331)	39(.331)	86(.729)	118	5.314	0.424
-	2 (.100)	15(.750)	17(.850)	20	5.300	0.014
2(.047)	14(.326)	15(.349)	31(.721)	43	5.172	0.128
-	9(.281)	15(.469)	24(.750)	32	5.062	0.110
-	2(.200)	5(.500)	7(.700)	10	4.600	0.462
1(.026)	11(.290)	12(.316)	24(.632)	38	4.476	0.124
2(.077)	4(.154)	10(.385)	16(.615)	26	4.312	0.164 → CUT-
-	3(.111)	12(.444)	15(.556)	27	3.552	0.760 OFF RANGE
-	1(.091)	5(.455)	6(.545)	11	3.458	0.094
2(.025)	15(.188)	22(.275)	39(.488)	80	3.404	0.054
-	22(.130)	47(.278)	69(.408)	169	2.708	0.696
-	1(.036)	11(.393)	12(.429)	28	2.646	0.062
-	2(.054)	3(.231)	5(.385)	13	2.618	0.028
-	5(.102)	13(.265)	18(.367)	49	2.406	0.212
-	4(.148)	5(.185)	9(.333)	27	2.294	0.112
-	-	4(.364)	4(.364)	11	2.184	0.110
-	5(.102)	10(.204)	15(.306)	49	2.040	0.144
-	1(.042)	4(.167)	5(.208)	24	1.338	0.702
-	1(.003)	3(.130)	4(.174)	23	1.124	0.214

IS: The figures in brackets represent the proportions.

If p_1 , p_2 , p_3 are the proportions in I, II, III divisions respectively, then $A.Q. = 10p_1 + 6p_2 + 6p_3$.

APPENDIX II A

Formulae for getting the Variables

No.	Variable	Data	Formula
	Total no. of hours for instruction in the academic year: x_1	1. Actual number of teaching days (annual) : D 2. Number of working hours per day: H	$x_1 = DH$
	Size of the section in school leaving class: x_2	1. Enrolment in class XI : E 2. Number of sections in Class XI : S	$x_2 = \frac{E}{S}$
	Pupil-teacher ratio in secondary classes: x_3	1. Enrolment in Secondary classes : E_1 2. Total number of teachers in secondary classes: T	$x_3 = \frac{E_1}{T}$
	Book value of equipment per pupil: x_4	1. Total book value of equipment in Rs.: V 2. Total enrolment of the school: N	$x_4 = \frac{V}{N}$
	Annual recurring expenditure per pupil : x_5	1. Total annual recurring expenditure : R 2. Total enrolment of the school : N	$x_5 = \frac{R}{N}$
	Number of library books per pupil: x_6	1. Total no. of library Books : B 2. Total enrolment of the school: N	$x_6 = \frac{B}{N}$

Measurements on Factors

Factors						
S.No.	x_1	x_2	x_3	x_4	x_5	x_6
HIGH ACHIEVERS						
1.	892	34	45	111.96	155.70	9.09
2.	1050	24	25	41.58	209.45	10.91
3.	1100	21	19	42.60	273.18	9.03
4.	915	35	23	64.92	163.78	9.76
5.	850	28	21	42.46	145.33	2.22
6.	1266	27	23	174.12	187.15	10.01
7.	988	27	26	24.79	203.36	5.08
8.	1302	29	20	208.46	241.76	15.65
9.	835	38	22	60.90	183.45	5.87
10.	950	39	27	35.00	179.24	5.57
11.	850	20	17	112.50	226.67	4.40
12.	880	22	36	62.76	163.26	4.33
LOW ACHIEVERS						
1.	865	27	27	15.17	102.55	3.03
2.	885	11	15	50.00	152.98	4.08
3.	985	28	23	38.37	152.53	7.21
4.	785	43	25	75.04	150.77	4.03
5.	1100	22	10	37.25	180.18	2.95
6.	1095	13	15	96.48	288.84	5.71
7.	765	23	28	81.89	255.52	4.77
8.	1065	14	15	52.05	186.08	4.65
9.	900	11	37	32.22	124.68	2.79
10.	1005	24	24	31.86	155.04	1.42
11.	700	24	20	45.23	186.61	4.29
12.	1070	12	20	57.75	122.26	7.19

APPENDIX III

**Differences in Mean Values of Factors between
High and Low Achieving Groups**

S.No.	Variable	\bar{X}_H	\bar{X}_L	$d = \bar{X}_H - \bar{X}_L$
1	X_1	989.8333	935.0000	54.8333
2	X_2	28.6042	21.0000	7.6042
3.	X_3	25.3333	21.5833	3.7500
4.	X_4	81.8375	51.1075	30.7300
5.	X_5	194.3608	169.0033	25.3575
6.	X_6	7.6600	4.3433	3.3167

NOTES: \bar{X}_H and \bar{X}_L are the means of high and low achieving groups respectively.

Mean, $\bar{X} = \frac{\sum X}{N}$, where $\sum X$ is the total and N is the number of schools.

APPENDIX IV

Pooled Variances and Co-variances of Factors

Variable	X_1	X_2	X_3	X_4	X_5	X_6
X_1	22362.6123	-404.5564	-424.6970	2950.4593	2164.3919	256.2795
X_2		67.3768	13.2002	-13.6455	-99.1726	-0.4143
X_3			58.1629	-33.6815	-158.1685	-2.2097
X_4				1985.3606	64.4029	77.9532
X_5					1996.1401	40.6990
X_6						8.5026

NOTES: Pooled variance (X_1) = $\frac{\sum X_{1H}^2 + \sum X_{1L}^2 - \left[\left(\frac{\sum X_{1H}}{N_H} \right)^2 + \left(\frac{\sum X_{1L}}{N_L} \right)^2 \right]}{N_H + N_L - 2}$

Pooled co-variance (X_i, X_j) = $\frac{\sum X_{iH} X_{jH} + \sum X_{iL} X_{jL} - \left[\left(\frac{\sum X_{iH}}{N_H} \right) \left(\frac{\sum X_{jH}}{N_H} \right) + \left(\frac{\sum X_{iL}}{N_L} \right) \left(\frac{\sum X_{jL}}{N_L} \right) \right]}{N_H + N_L - 2}$

Where X_{iH} , X_{iL} are measurements on X_i in high and low achieving groups respectively and N_H , N_L are number of schools.

APPENDIX V

Pivotal Condensation Method for obtaining successive 'D²' values and the coefficients of discriminant function at each stage.

Row No.	x_1	x_4	x_5	x_6	x_2	x_3	Differences in means
01.	22362.6123	2950.4593	2164.3919	256.2795	-404.5564	-424.6970	54.8333
02.		1985.3606	674.4029	77.9532	-13.6455	-33.6814	30.7300
03.			1996.1401	40.6990	-99.1726	-158.1685	25.3575
04.				8.5026	-0.4143	-2.2097	3.3167
05.					67.3768	13.2002	8.6042
06.						58.1629	3.7500
07.							0.0000

Operation I

10.	1	0.131937	0.096786	0.11460	-0.018091	-0.018991	0.002452
11.	0.131937	1596.085851	388.839746	44.140936	39.731259	22.350673	23.495474
12.	0.096786		1786.657266	15.895069	-60.016586	-17.064533	20.050411
13.	0.114600			5.565637	4.222052	2.657304	2.686363
14.	-0.018091				60.057970	5.517269	8.596172
15.	-0.018991					50.097479	4.791357
16.	0.002452						-0.134457 =-D ₁ ²

Operation II

20.	0.000083	1	0.243621	0.027656	0.024893	0.014003	0.014721
21.	0.064512	0.243621	1691.927738	5.141317	-69.695974	-122.509456	14.326301
22.	0.007796	0.027656		4.344875	3.123252	2.039198	2.038504
23.	-0.024389	0.024893			59.068940	4.960912	8.011288
24.	-0.020846	0.014003				49.784503	4.462333
25.	0.000502	0.014721					-0.480328 =-D ₂ ²

Operation III

30.	0.000038	0.000144	1	0.003039	-0.041193	-0.072408	0.008467
31.	0.007601	0.026916	0.003039	4.329251	3.335038	2.411470	1.994972
32.	-0.018741	0.034929	-0.041193		56.197954	-0.085634	8.601404
33.	-0.016191	0.031644	-0.072408			40.913838	5.499621
34.	-0.000042	0.012658	0.008467				-0.601629 =-D ₃ ²

contd....

APPENDIX V contd.

Row No.	x_1	x_4	x_5	x_6	x_2	x_3	Differences in means
Operation IV							
40.	0.001756	0.006217	0.000702	1	0.770350	0.557018	0.460812
41.	-0.024597	0.014195	-0.043534	0.770350	53.628807	-1.943310	7.064578
42.	-0.020426	0.016652	-0.074101	0.557018	39.570606	4.388387	
43. ^(*)	-0.003545	0.000255	0.007067	0.460812		-1.520936 ₂	
							= - D ₄
Operation V							
50.	-0.000459	0.000265	-0.000812	0.014364	1	-0.036236	0.131731
51.	-0.021318	0.017167	-0.075679	0.584932	-0.036236	39.500188	4.644381
52. ^(*)	5.000302	-0.001617	0.012803	0.359336	0.131731	-2.451560 ₂	
							= - D ₅
Operation VI							
60.	-0.000540	0.000435	-0.001916	0.014808	-0.000917	1	0.117579
61. ^(*)	0.002206	-0.003637	0.021702	0.290562	0.135990	0.117579	-2.997642
							= - D ₆

NOTES:

Operation I:

Row 10 = Row 01 + 22362.6123

Row 11 = Row 02 - 2950.4593 X Row 10.

Row 12 = Row 03 - 2164.3919 X Row 10.

Row 13 = Row 04 - 256.2795 X Row 10.

Row 14 = Row 05 - (-404.5564) X Row 10.

Row 15 = Row 06 - (-424.6970) X Row 10.

Row 16 = Row 07 - 54.8333 X Row 10.

(The figure in row 10, coming after 1 are written in the column under 1, with indented marks, and operation V is carried on)**Operation II:**

Row 20 = Row 11 ÷ 1596.085851

Row 21 = Row 12 - 388.839746 X Row 20 etc.

Similarly the remaining operations can be performed.

Checks provided:

1. Sum check (not shown in the table)

2. At every stage $\sum l_i^2 d_i^2 = D_i^2$

APPENDIX VI

'Z' Values for the High and Low Achieving Schools

S.No.	High Achievers (Z_H)	Low Achievers (Z_L)
1.	7961.68	5351.46
2.	7290.94	4322.53
3.	7213.86	[6323.91]
4.	7069.81	[6658.57]
5.	[5316.66]	5088.90
6.	7013.42	6130.43
7.	6667.05	[6387.10]
8.	8251.72	5053.96
9.	6827.60	5091.11
10.	7252.86	5454.28
11.	[5612.98]	5571.78
12.	[6197.12]	4930.31

NOTES: $\bar{Z}_H = 6889.64$ $\bar{Z}_L = 5530.36$

$$\bar{Z} = 6210.00$$

The schools with indented values would have got misclassified, by using this discriminant function.